AD_____

Award Number: W81XWH-11-C-0004

TITLE: Objective Assessment of Post-Traumatic Stress Disorder Using Speech Analysis in Telepsychiatry

PRINCIPAL INVESTIGATOR: Pablo Garcia

CONTRACTING ORGANIZATION: SRI International, Menlo Park, CA 94025

REPORT DATE: December 2012

TYPE OF REPORT: Annual Report

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
December -2012	Annual	29 October 2011- 28 November 2012

4. TITLE AND SUBTITLE
Objective Assessment of Post-Traumatic Sress Disorder Using Speech Analysis in Telepsychiatry

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-11-C-0004

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Pablo Garcia and Bruce Knoth

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
SRI International
Menlo Park, CA 94025

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release;
Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

(The abstract in Block 14 **must** state the purpose, scope, major findings and be an **up-to-date** report of the progress in terms of results and significance.)
The objective of this project is to explore the feasibility of using speech features to assess the Post-Traumatic Stress Disorder (PTSD) status of a patient. The premise for this project is that an individual's speech features, drawn from a recorded CAPS interview, correlate to the diagnosis of PTSD for that person. Recorded interviews from a patient population will be used to develop and test an objective scoring system. NYU is collecting speech data from patients and supplying it to SRI International. SRI now has data from 20 PTSD-negative patients and 13 PTSD-positive patients. Few PTSD-positive patients meet the inclusion criteria for the study so it is taking much longer than expected to acquire the target of 20 PTSD-positive samples. Preliminary tests using this small dataset show promise in predicting PTSD based on speech characteristics.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF: Unclassified			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Mary Kelly
Unclassified	Unclassified	Unclassified		10	19b. TELEPHONE NUMBER *(include area code)* 650-859-5950

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

CONTENTS

1. INTRODUCTION

SRI International (SRI) is pleased to provide this annual report for the *Objective Assessment of PTSD Using Speech Analysis in Telepsychiatry* project, contract number W81XWH-11-C-0004, covering the period 01 January 2012 – 31 December 2012. The objective of this project is to explore the feasibility of using speech features to assess the Post-Traumatic Stress Disorder (PTSD) status of a patient. The premise for this project is that an individual's speech features, drawn from a recorded Counseling and Psychological Services (CAPS) interview, correlate to the diagnosis of PTSD for that person. Recorded interviews from a patient population will be used to develop and test an objective scoring system.

2. TASKS

TASK 1: DEVELOP THE STUDY PROTOCOL AND SUBMIT IT TO THE APPROPRIATE INSTITUTIONAL REVIEW BOARDS (NYU/SRI).

Task Description: SRI and NYU will develop a protocol to select, prepare, and analyze recorded interviews from a patient population screened for PTSD. The population will include both PTSD-negative and PTSD-positive patients. The protocol will include appropriate informed-consent procedures and procedures for de-identifying the data to eliminate the 18 Health Insurance Portability and Accountability Act (HIPAA) identifiers (45 C.F.R. § 164.514(b)(2)(i)(A) – (R)). The protocol will be submitted to IRBs at NYU, SRI, and the United States Army Medical Research and Materiel Command (USAMRMC) for approval.

Progress: In August 2011, SRI received a determination that this project does not require further review (HRPO Log Number A-16207). SRI forwarded the determination to NYU. This task is complete.

TASK 2: PREPARE THE DATA FOR ANALYSIS IN CONFORMANCE WITH THE STUDY PROTOCOL (SRI/NYU).

Task Description: After IRB approval, NYU personnel will de-identify the data per the study protocol. Then, with assistance from SRI, they will transcribe the interviews and segment the recordings into interviewer and interviewee units. The resulting data will be provided to SRI.

Progress: NYU has now collected data from 33 patients (20 are PTSD-negative and 13 are PTSD-positive) and transferred these files to SRI in an encrypted format. Three early recordings of PTSD-negative patients were removed from the study due to poor audio quality and are not included in the 33 recordings. Every subject who has met the inclusion criteria has been male except for one. NYU and SRI decided to remove the one PTSD-negative recording of a female from the study to eliminate gender influence from the dataset. The eligibility criteria for this

study are rigorous, and NYU is collecting data from about one PTSD-positive patient per month, so this data collection process is proceeding slowly. The goal is to collect data from a total of 40 patients, split evenly between PTSD-positive and PTSD-negative diagnoses. NYU doesn't know whether a subject is PTSD-positive or negative until after the subject has been recruited and tested, so it is not possible to recruit specifically for one group or the other. So far, most of the subjects who have consented to the study have been PTSD-negative. The required number of PTSD-negative samples have now been collected.

SRI received a no-cost extension for this contract through December 2013.

TASK 3: DEFINE AND EXTRACT PROSODIC FEATURES FROM THE DATA, RUN AUTOMATED SPEECH RECOGNITION (SRI/NYU).

Task Description: SRI, with assistance from NYU, will define and extract prosodic features from the interviewee's recording segments created in Task 2. These features include parameters such as phonetic and pause durations and measurements of pitch and energy over various extraction regions. Automated speech recognition will be used to transcribe these segments.

Progress: SRI has received 33 interviews from NYU, 13 of which are PTSD positive. All the recordings have been manually segmented to delineate the sections of the recordings where patients are speaking. Pitch, energy, and spectral tilt features have been extracted from 28 of the recordings and are being used to investigate classifying PTSD-positive patients vs. PTSD-negative patients based on these speech characteristics. Features from the remaining five subjects will be included in the analysis shortly.

TASK 4: EXTRACT LEXICAL FEATURES FROM TRANSCRIPTS (SRI/NYU).

Task Description: SRI will extract lexical features from the interviewee transcripts created in Task 2. Features may include disfluencies, idea density, referential activity, analysis of sentiment, topic modeling, and semantic coherence.

Progress: This task has not yet started.

TASK 5: TRAIN THE STATISTICAL MODEL USING MACHINE-LEARNING ALGORITHMS (SRI).

Task Description: Using the outputs from Tasks 3 and 4, SRI will perform feature selection via univariate analysis and apply machine-learning algorithms to develop models that predict outcome measures, such as PTSD status, and aspects of the CAPS scores on the basis of acoustic and lexical feature inputs.

Progress: We have performed initial experiments to identify PTSD-positive patients and PTSD-negative patients using mel frequency cepstral coefficients and prosodic polynomial coefficients. We continue to update the experiments as new recordings are received. These standard features are used in many speech classification protocols based on Gaussian mixture models (GMMs). We also applied universal background models (UBMs) based on the same cepstral or polynomial coefficients so that we can use the joint factor analysis (JFA) modeling approach. These UBMs were developed from data previously used by SRI for speaker identification.

TASK 6: VALIDATE MODEL AND ANALYZE RESULTS (SRI).

Task Description: SRI will validate the PTSD assessment model and measure its reliability using statistical analysis techniques, such as N-fold cross-validation and split-half reliability.

Progress: SRI has tested classifiers based on acoustic features, prosodic features, and a fusion of the both acoustic and prosodic features. Although we now have data from ten PTSD-positive patients, these results are based on eight subjects (we have not yet rerun the calculations with the ten patients).

The classifiers' accuracy was tested using N-fold leave-one-out cross-validation. In this framework, if we have N training samples, the model is trained on N-1 samples and tested on the held-out sample. This process is iterated N times, leaving out a different sample each time. The final accuracy is the cumulative result across all N samples.

In our prior quarterly report, we had reported accuracies of 62% - 87%, depending on which feature set (acoustic or prosodic) and data group was used. Table 1 shows these reported results. These results aimed to demonstrate the best achievable accuracy in the target group. The features and decision thresholds were optimized on the whole set of recordings to achieve the reported accuracies. We believe these results are very important, since they demonstrate the discriminative potential of the features we are using, but because of the very limited number of speaker samples available for this study, these results may not generalize to a larger population.

Table 1: Previously Reported Preliminary Results (Best Case)

- Average N-fold accuracy on whole data

	8 PTSD+ vs. 20 PTSD-	8 PTSD+ vs. 8 PTSD- (Group 1)	8 PTSD+ vs. 8 PTSD- (Group 2)
Majority	71.4%	50.0%	50%
Acoustic	71.4%	75.0%	81.3%
Prosodic	82.1%	81.3%	81.3%
Fusion	78.6%	81.3%	87.5%

- Average N-fold accuracy for Military Trauma Section

System	8 PTSD+ vs. 20 PTSD-	8 PTSD+ vs. 8 PTSD- (Group 1)	8 PTSD+ vs. 8 PTSD- (Group 2)
Majority	71.4%	50.0%	50%
Acoustic	71.4%	68.8%	75%
Prosodic	78.6%	81.3%	87.5%
Fusion	82.1%	81.3%	93.8%

We have since analyzed the data using a more conservative approach, to avoid possibly over-fitting to the limited data.

Figure 1 shows the modules in the machine-learning training system. Input audio is processed by the Feature Extraction Module. It computes thousands of parameters, or "features," from the audio data, and identifies the more representative features to use for classification purposes. These features are used by GMMs that comprise the two target class models (one for PTSD-positive and one for PTSD-negative). Each of these two models generates a score, given the features for a given audio set. The scores are converted to posterior probabilities and the ratio of the posterior probability of PTSD+ over the posterior probability of PTSD- is computed. If the ratio is above a specified threshold, the subject is classified as PTSD-positive; otherwise the subject is classified as PTSD-negative.

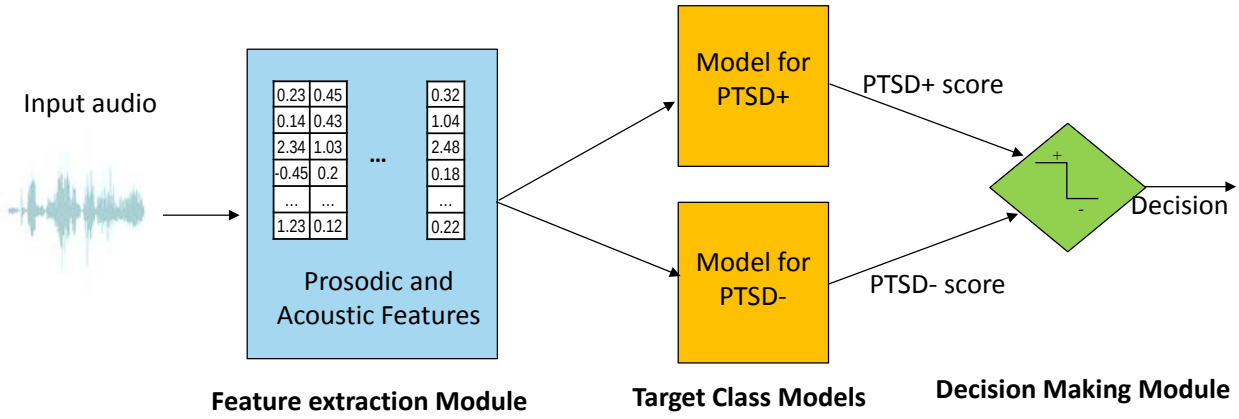

Figure 1. Modules comprising the trainer.

There are three general opportunities to over-fit the algorithm to a given set of data. The first is in choosing the features to use (one subset of features may be able to discriminate among two speaker groups more accurately than any other feature set). The second over-fitting opportunity is in training the GMM classifier (when using more mixtures, the model may fit the training data better, but won't generalize). The third opportunity is the threshold level used by the decision-making module (the threshold needs to be chosen on a held-out set, representative of the test population). The GMM classifier was always trained fairly, since we train and test it with an N-fold, leave-one-out process, which chooses a size that doesn't over-fit the data. But the other two, the feature and threshold selections, were optimized on the whole dataset for the results presented in Table 1 and may be over-fit.

We have now taken a more conservative approach and re-analyzed the data. We made two major modifications in our analytical procedure. First, rather than choose the subset of features that gave the best results for our PTSD speech data, we used features that have independently been shown to be highly effective for speaker identification. This selection may be too conservative, because features that are effective for speaker identification may not be most useful in generating psychological measures, such as PTSD or depression classifications.

Second, rather than treating each speaker as a single sample point (extracting a single set of features from a speaker), we split the speech into shorter segments. We extract a feature vector from each of these segments and treat it as a training sample. This way, we have many more samples to input into the statistical learning algorithms, which results in more robust models. We experimented with segments with length of 30, 60, and 90 seconds.

Third, rather than select one threshold for the decision-making module, we assess system accuracy across the full range of thresholds and present those results in a ROC curve (Figure 2). Figure 2 shows a graph with four curves. The ordinate of this plot represents the false-negative rate and the abscissa represents the false-positive rate. One of the four curves is a straight line through the center of the graph. This line represents a classifier that randomly guesses if any given sample is positive or negative. The line spans from the extreme of guessing that every sample is negative (resulting in a 100% false-negative rate) to the other extreme of assigning every sample to the positive category. At the mid-point, it designates half the samples as positive and half as negative, resulting in 50% false-positive and false-negative rates (assuming equal numbers of true-positive and true-negative samples).

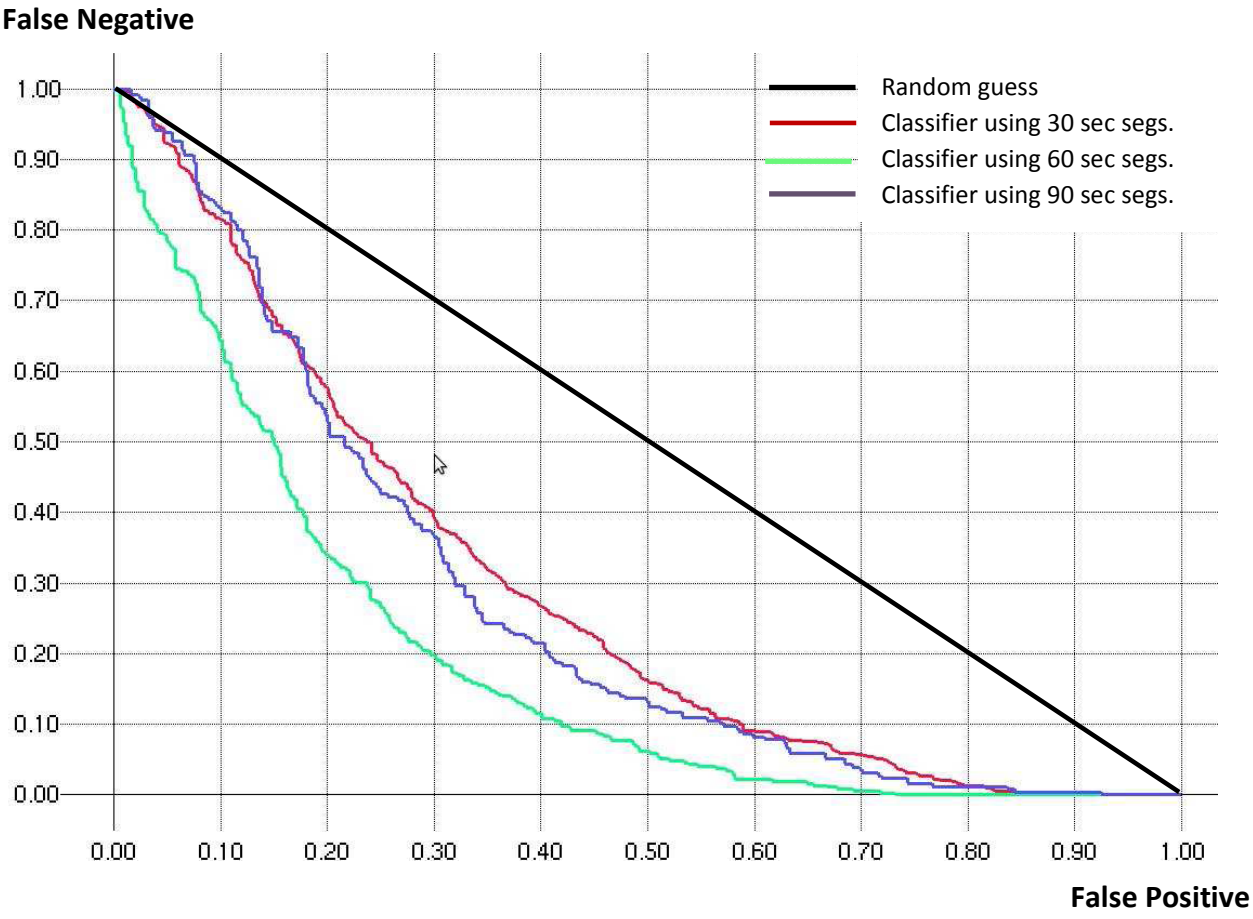

Figure 2. Classifier performance.

The other three curves represent results from our classifier based on acoustic features. These three curves differ only in the length of each sample (one curve represents the recordings broken into 30-second segments, and the other two represent 60-second and 90-second segments). The

plot shows the best results using the 60-second segments, with roughly a 25% false-positive and false-negative rate at the mid-point (the other two curves have lowest rates of about 33%). These results are based on features optimized for speaker identification, not for PTSD, and the size of the GMM model is one (we trained a single Gaussian for each class), so the results may be a conservative representation of the potential of our approach.

Although the model parameters are always trained using data separated from the held-out test sample, the results we report are the "best-case scenario," since we report the results of the best possible model configuration for each experiment (among 90 different configurations for the prosodic coefficients and 36 for the MFCC features). We also choose the decision threshold for each experiment so as to optimize the accuracy for this test data. Our results show that there is a model configuration and decision point with these features that makes the two classes (PTSD-positive and -negative) separable – better than the guessing using the majority rule. Although this particular model configuration and threshold may not apply to much larger datasets collected from multiple sources, these results show promise for using speech as a predictor of PTSD status.

3. KEY RESEARCH ACCOMPLISHMENTS

There are no key research accomplishments because the project is in an early stage of data collection.

4. REPORTABLE OUTCOMES

Not applicable at this time.

5. CONCLUSIONS

Not applicable at this time.

6. REFERENCES

Not applicable at this time.

7. APPENDICES

None